# SCALABLE COMPRESSION OF 3D MEDICAL DATASETS USING A (2D+T) WAVELET VIDEO CODING SCHEME

*Marion Bénetière, Vincent Bottreau, Antoine Collet-Billon and Thomas Deschamps*
Laboratoires d'Electronique PHILIPS
22, avenue Descartes, B.P. 15, F-94453 LIMEIL-BRÉVANNES Cédex FRANCE
m.benetiere@philips.com, bottreau@philips.com, antoine.collet-billon@philips.com, thomas.deschamps@philips.com

## ABSTRACT

Nowadays the all digital solution in hospitals is becoming widespread. The formidable increase of medical data amount to be processed, transmitted and stored, requires some efficient compression systems and innovative tools to improve data access, while ensuring a sufficient visualization quality for diagnosis. Medical image sequences can benefit from advanced video coding techniques when adapted to their specific constraints. *Scalability,* or the capability to partly decode a video bitstream and to get a reconstruction quality proportional to the decoded amount of information, is a key functionality.

We have developed a video codec based on a 3D Motion-Compensated subband decomposition, which provides a combination of temporal, spatial and SNR scalabilities together with a very competitive compression ratio. We show that, when applied to medical sequences, it outperforms JPEG-2000 on coding efficiency aspects and offers new functionalities.

## 1. INTRODUCTION

In hospitals, more and more radiology departments are becoming all digital, which leads to a formidable amount of digital data (several terabytes per year in cardiovascular department alone). Also, the progresses in the different imaging modalities will lead to always more data per exam. For instance in X-ray cardiology the future flat solid state dynamic detectors will deliver up to thirty 8 Mbytes image per second. Even though storage cost is decreasing, transmission time and archival cost could benefit from more efficient compression techniques. In parallel there is also a development of telemedicine, where this requirement is even more important, as is the need for efficient tools to browse databases. Consequently, compression systems are expected to become flexible or *scalable*, that is to be able to adapt a single video bitstream to variable transport conditions (bandwidth, error rate…) and to varying receiver capabilities and demands (display size, preview or diagnosis application…). Scalability consists in delivering compressed information in an embedded way, that could be progressively decoded (or sent) to get a reconstruction quality (also size or frequency for video sequences) proportional to the amount of decoded (or received) data.

In medical imaging, the new DICOM standard [1] for lossy compression is JPEG-2000 [2], which replaces JPEG and overcomes some of its limitations (compression at low bit-rates i.e. below 0.25 bpp or transmission in noisy environments). That is a fully *scalable* (in spatial resolution and quality) still picture compression system, based on a subband decomposition and an efficient wavelet coefficient coding technique called *Embedded Block Coding with Optimized Truncation* (EBCOT) [3], which answer the need for increasing resolution and fidelity by allowing a fully progressive transmission. However the achievable compression ratio with such an intra-image compression scheme is limited, for instance 20:1 at best for MR images. Above, the alteration of the image content becomes unacceptable for diagnosis, as shown by careful clinical evaluations [4].

The only way to further improve coding efficiency on medical sequences is to better exploit the temporal (or z-axis) redundancies between the successive images of 3D stacks (as if they were video data) by using an inter-frame scheme, but without giving up the high level of scalability offered by JPEG-2000[1]. In theory, introducing the temporal dimension opens the way to temporal scalability and related applications.

In current standards for video compression, like MPEG-4, which are based on block DCT coding of Displaced Frame Differences (DFD), scalability is implemented through additional layers of the single-scale prediction loop that delivers one *base* and one (or more) *enhancement* video bitstream(s) (usually two spatial or temporal resolutions are then available). However the proposed solutions have a very limited efficiency. Above all, the combination of the three types of scalability is still an issue in current standards, although some contributions have recently appeared [5].

The extension of the 2D to a 3D (or 2D +t) wavelet analysis for video by including the temporal dimension within the decomposition appears as a promising framework for medical applications. Indeed the coefficient coding strategies used in the 2D case like the Embedded

[1] Motion JPEG-2000, the future extension to JPEG-2000 is not a good candidate, because it remains an intra-scheme.

Zerotrees Wavelet (EZW) [6] or Set Partitioning In Hierarchical Trees (SPIHT) [7] algorithms, which unlike EBCOT, take advantage of the dependencies existing along hierarchical spatiotemporal trees, were successfully adapted to video (3D SPIHT [8]), yielding very competitive compression performances, together with desirable properties of bitstream embeddedness. Since such approaches were only tailored for SNR-scalability, we have developed a new approach, which extends these coding principles to a fully scalable video coding scheme.

We will show that our 3D codec provides a valuable solution for video compression, transmission and manipulation in the medical field and outperforms the JPEG-2000 standard both in terms of compression efficiency and scalability properties.

The article is organized as follows: Section 2 gives an overview of the 3D codec. Section 3 presents the coding method and bitstream organization for full scalability. Section 4 highlights the added value of such a scalable compression scheme for medical applications. Simulation results on medical sequences are shown in Section 5 and Section 6 gives some concluding remarks.

# 2. OVERVIEW OF THE 3D CODEC

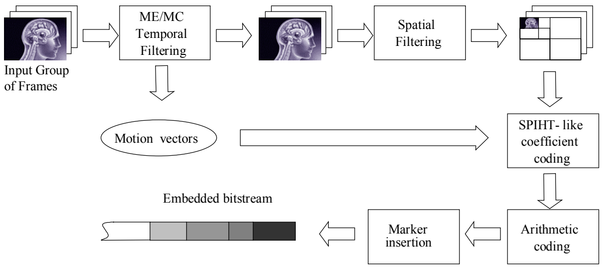



*Figure 1: Overview of the encoder structure.*

The 3D encoder structure is organized as shown in Figure 1. The decoder is fully symmetric. The proposed codec is based on a 3D wavelet analysis, consisting of Motion-Compensated Temporal Filtering (MCTF) and Spatial Filtering, which leads to a spatiotemporal multiresolution decomposition of the input Group Of Frames (GOF). Thus, lower display sizes and/or lower framerates may naturally be obtained.

## 2.1. Motion-Compensated Temporal Filtering

In this 3D subband decomposition scheme, the input video is first temporally filtered as shown in Figure 2. Each frame is considered as a temporal tap, leading to temporal subbands containing several frames. MCTF temporally filters the GOF in the motion direction [9] and results in a temporal decomposition tree in which the leaves (temporal subbands) contain several frames. These frames are further spatially decomposed and yield the spatiotemporal subtrees of wavelet coefficients.

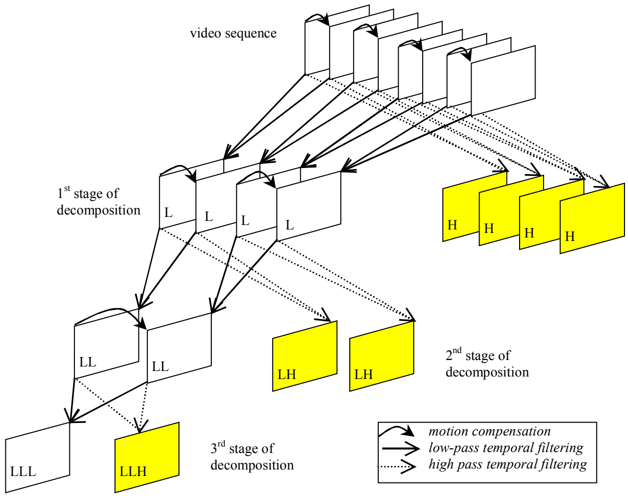



***Figure 2: Temporal multiresolution analysis with motion compensation.***

### 2.1.1. Temporal Filtering

In our approach Haar filters were used for TF on a GOF of 16 frames, which is a good trade-off between delay and energy. With Haar filters, motion estimation (ME) and motion compensation (MC) are only performed every two frames of the input sequence and the total number of ME/MC operations required for the whole temporal tree is roughly the same as in a predictive scheme (see Figure 2).

## 2.2. Spatial Decomposition

Concerning the spatial wavelet analysis, there are fewer restrictions in the choice of the filter bank than for the temporal decomposition. However, the SPIHT algorithm relies on the assumption of orthonormal decomposition. Biorthogonal transforms may be employed with an appropriate renormalization of the energy in the subbands. We implemented and tested a battery of filter banks. The filter bank choice is encoded and sent as side information in the bitstream. Implementation is made in a very simple way using the lifting technique (or *ladder* scheme) [10].

# 3. CODING METHOD

## 3.1. Principle

Inspired from the SPIHT [7][8], our algorithm looks for zero-trees in the wavelet subbands in order to reduce redundancies between them. The wavelet coefficients are encoded according to their nature: *root* of a possible zero-tree or *insignificant set*, *insignifican*t pixel and *significant* pixel.

In this approach, significance map and bitplane coding are combined. The significance map is efficiently encoded by exploiting the inter-subband correlations and the bitplane approach is retained to encode the refinement bits.

An iterative process of in-depth search successively scans and encodes the coefficients of each spatiotemporal tree (see Figure 3).

To preserve the initial subband structure, we replaced the list management (one list per type of coefficients) of the original SPIHT algorithm, by a subband scanning and a flag interpretation. A flag is added to each coefficient in order to indicate which type of encoding will be performed on this coefficient: set-significance and/or pixel-significance.

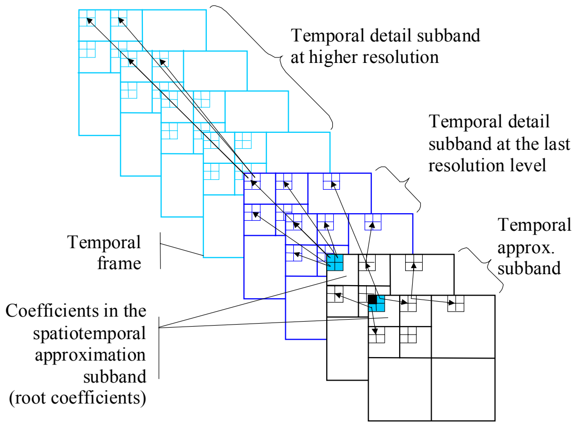



*Figure 3: Examples of parent-offspring dependencies in the spatial temporal orientation tree (3D case).*

From the maximum significance level down to 0, a full exploration of the spatiotemporal subbands is performed in an order that respects the parent-offspring relationships. Flags are updated for each coefficient according to the original SPIHT *pixel and set significance* rules.

### 3.2. Scanning order

This algorithm is very flexible and provides several progressive scanning modes. Indeed, the spatiotemporal volume of coefficients can be explored either by following its temporal, spatial or "diagonal" orientation. Thus three types of "multi-scalable" bitstreams may be obtained, one led by spatial resolution, the second by the temporal resolution and the third being an hybrid version of the two resolutions.

Here, we have favored the temporal scalability. For each bitplane, tree scanning is temporally oriented, since in this scheme the temporal resolutions are fully explored one after the other as shown in Figure 4. Inside each temporal scale all spatial resolutions are successively scanned and therefore all spatial frequencies are available.

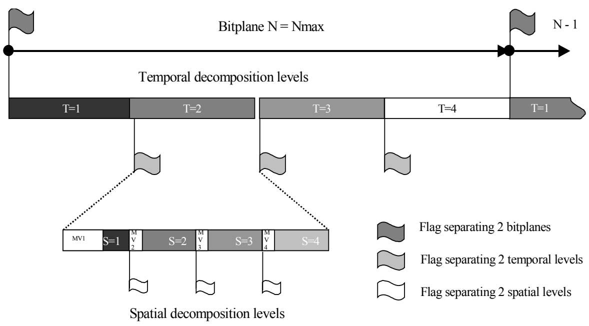



*Figure 4: Bitstream organization when temporal ordered is favored.*

Inside each spatiotemporal subband, coefficients are scanned horizontally or vertically according to the direction of the details in the subband. This particular scanning order also facilitates the context determination used to encode the current pixel by arithmetic coding with the Context Tree Weighting technique [11]. The evaluated contexts are all the more pertinent given that the proposed *SPIHT*-like coding strategy preserves the pixel neighborhoods within subbands and therefore their coherence.

### 3.3. Marker coding

If the wanted spatial resolution and/or framerate and/or quality are inferior to the ones provided by the encoder, the decoder (or server) must skip some parts of the bitstream. It is achieved by the introduction of special flags that indicate the end of a spatial level, a temporal level, a bit plane level or the end of a GOF. Due to the above-mentioned strategy, scalability is inherently associated to the bitstream organization and directly available at the decoder (or server) side, without any further processing.

## 4. ADDED VALUE OF SCALABILITY FOR MEDICAL APPLICATIONS

We are convinced that telemedicine can greatly benefit from advanced tools initially developed for scalable video compression, when adapted to the specific needs of the medical field. If spatial and quality scalabilities are already widespread, typically for refreshing a progressively transmitted image, the combination of temporal, spatial and quality scalabilities associated with a competitive coding efficiency is more innovative and powerful. For instance, image sequences like MR or 3D CT stacks are encoded at a reasonable compression ratio with maximal image size and frame rate. When a physician wants to browse a patient file in order to choose the type of exam needed for the diagnosis, he asks the server for a degraded version of each sequence with minimal display size and frame-rate. The server sends a preliminary version of each original bitstream, which can be quickly transmitted, decoded and visualized. Once he has identified the relevant image sequence, the physician asks the server for the missing parts of the corresponding bitstream. These data may be sent to the workstation dedicated to diagnosis with full screen display, full frame rate and top reconstruction quality.

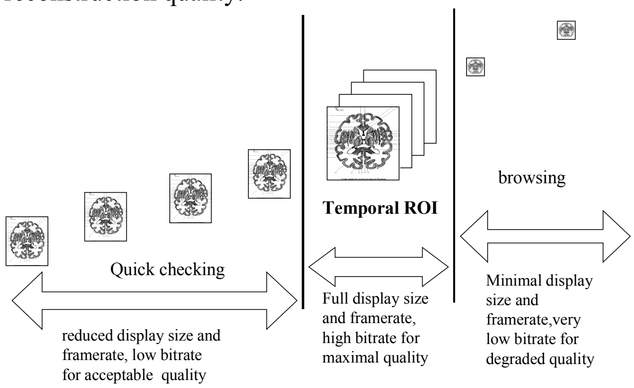
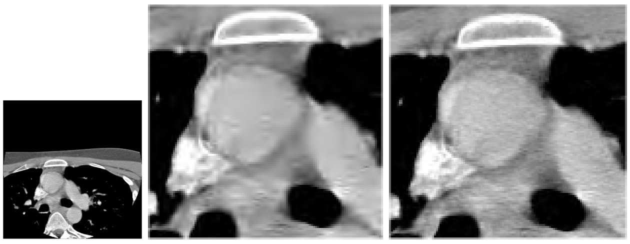



*Figure 5: Choice of a temporal Region of Interest (ROI) inside medical 3D stack.*

During their diagnosis, practicians often focus on a subset of images inside the sequence. To accelerate data access, the proposed system may give the possibility to choose a temporal Region of Interest (ROI) within the image sequence. Inside the temporal window, the data would be decoded with a higher bitrate at full resolution and frame-rate, whereas the rest of the sequence would be far more degraded, as shown in Figure 5.

## 5. EXPERIMENTAL RESULTS

The following illustrates the results obtained with a 3D CT data set and a 3D MR image sequence. They are compared with the JPEG2000 technique (C implementation available from [12]). Test parameters were chosen so as to lead to a fair comparison (irreversible wavelet transform 9/7, same number of decomposition levels, no spatial ROI and progression by layer). Moreover, motion compensation has been disabled since the translational assumption of our motion model does not suit displacements inside image stacks. However, temporal correlations are still taken into account by the spatiotemporal decomposition.



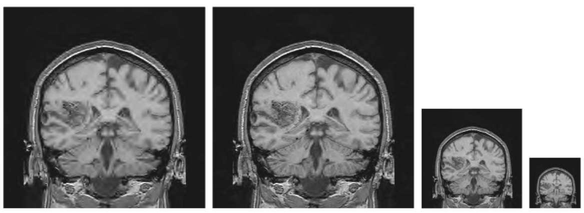
***(a) (b) (c)***

***Figure 6: (a) original 3D CT sequence image, reconstructed image at compression ratio 20 (b) zoom on JPEG2000, (c) zoom on 3D codec.***

Figure 6 shows a comparison between various reconstructed medical image sequences compressed at a compression ratio of 20 with the JPEG-2000 coder and our 3D codec. Figure 7 demonstrates spatial scalability on 3D MR sequence.

| | JPEG-2000 | 3D codec |
|---|---|---|
| 3D CT (224 images) | 36.83 dB | 38.53 dB |
| 3D MR (144 images) | 32.44 dB | 34.13 dB |

***Table 1: PSNR results on medical image sequences compressed with ratio 20.***



***(a) (b) (c) (d)***

***Figure 7: 3D MR sequence encoded at compression ratio 20 (a) JPEG-2000; 3D codec at (b) full, (c) half and (d) quarter spatial resolution from the same bitstream.***

Note that a careful clinical validation (with expert review) is required, in order to precisely assess the acceptable compression ratios, which closely depends on the medical application. The PSNR is just given as a coarse quality indicator (see Table 1).

## 6. CONCLUSION AND FUTURE WORKS

We successfully adapted a fully scalable video compression system to medical image sequences. Using this approach, groups of successive pictures are first temporally filtered and then spatially decomposed with wavelets. The spatiotemporal coefficients are further scanned and efficiently compressed using a new *SPIHT*-like strategy together with arithmetic encoding, which provides both competitive compression ratios and full bitstream embeddedness. By exploiting the temporal redundancies between successive images, this scheme outperforms JPEG-2000, the actual DICOM standard for still-pictures. The combination of the three types of scalability also offers new perspectives for telemedicine towards improved tools of database browsing or remote diagnosis.

## 7. REFERENCES


[1] Digital Imaging Communication in Medicine (DICOM), Final Draft 2000, available on web site, http://medical.nema.org/dicom.html/

[2] JPEG 2000 Part I Final Draft International Standard, ISO/IEC JTC1/SC29 WG1, JPEG 2000, available on official JPEG web site, http://www.jpeg.org/.

[3] D. Taubman, "High Performance Scalable Image Compression with EBCOT", *IEEE Trans. On Image Processing*, vol. 9, no. 7, pp. 1158-70, July 2000.

[4] S. Terae et al, "Wavelet Compression on Detection of Brain Lesions With Magnetic Resonance Imaging", *Journal of digital imaging*, Vol. 13, no. 4, November 2000, pp. 178-180.

[5] Mihaela van der Schaar, "All FGS temporal-SNR-spatial scalability", *Contribution to 54th MPEG Meeting, m6490*, La Baule France, Oct. 2000.

[6] J. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3445-3462, Dec. 1993.

[7] A. Said and W.A. Pearlman, "A New, Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, pp. 243-250, June 1996.

[8] Kim, Z. Xiong, W.A. Pearlman, "Low Bit-Rate Embedded Video Coding with 3D Set Partitioning in Hierarchical Trees (3D SPIHT)", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1374-87, Dec. 2000.

[9] J.-R. Ohm, "Three Dimensional Subband Coding with Motion Compensation", *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 559-571, Sept. 1994.

[10] A.R. Calderbank, I. Daubechies, W. Sweldens and B.L. Yeo, "Wavelet Transforms that Map Integers to Integers", *Applied Computational Harmonic Analysis*, vol. 5, no. 3, pp. 332-369, 1998.

[11] B. Felts and B. Pesquet, "Efficient Context Modeling in scalable 3D wavelet-based video compression", *Proc. of IEEE Int. Conf. on Image Processing*, pp. 1004-1007, Vancouver, Canada, Sept. 10-13, 2000.

[12] JasPer: A Software-Based JPEG-2000 Codec Implementation, available at The Jasper Project Home Page, http://www.ece.ubc.ca/~mdadams/jasper/.